Representation of Clinical Problem Assessment Phrases in U.S. Family Practice Using Read Version 3.1 Terms: A Preliminary Study

Robert R. Hausam, M.D.*
Allen W. Hahn, D.V.M., Ph.D.
Medical Informatics Group, University of Missouri-Columbia, Columbia, MO
* currently at Department of Medical Informatics, University of Utah, Salt Lake City, UT

The Read Codes from the United Kingdom are a comprehensive clinical vocabulary, and one of the most likely candidates for adoption as a standard for use in Computer-Based Patient Record (CPR) systems. The new version 3.1 codes represent a major enhancement to the content and structure of the coding system, including incorporation of a new hierarchy and an explicit model for the use of qualifier terms. This is a preliminary study investigating the suitability of these codes for representing clinical problem assessment phrases in U.S. family practice.

Problem assessment phrases from outpatient progress notes were encoded into the equivalent Read terms. The problem assessment phrases were evaluated for complexity and clarity. The coded representations of the phrases were evaluated for clinical acceptability. A list of coding difficulties was compiled. The most common difficulties were (1) qualifier terms present but not allowable for that Read concept (24%), and (2) qualifier terms not present (20%). British spelling and abbreviation variants were noted, but were relatively insignificant. The Read codes appear to be suitable for use in U.S. primary care practice with fairly minor modifications, but further development is required to expand the content and structure of the model for qualifier terms.

INTRODUCTION

There is major interest at present in the U.S. and the rest of the world in developing and implementing computer-based patient record (CPR) systems. These systems are seen as potentially providing significant benefits in clinical care, including reduced costs and improved quality.[1] It is almost universally felt that a controlled clinical vocabulary is an essential component of a CPR system. Use of a controlled vocabulary for data recording will facilitate communication between providers, enhance data retrieval for support of clinical care, quality assurance, and research, and especially will enhance the ability to incorporate automated decision support technology into CPR systems.[1,2]

Several comprehensive controlled clinical vocabularies have been and are being developed which may be candidates for adoption as standards for future CPR systems. The most likely candidates at present are SNOMED International, the Read Codes, and the UMLS.[1,3]

There have been few studies reported previously evaluating the Read Codes. One recent study sponsored by CPRI (Computer-Based Patient Record Institute) evaluated multiple vocabularies, including SNOMED International, the Read Codes (version 2), the UMLS Metathesaurus (version 1.3), and others.[3] SNOMED had the best overall performance, with UMLS and Read performing somewhat more poorly.

The Read Codes are a comprehensive clinical vocabulary and coding system originally developed by Dr. James Read, a general practitioner in the United Kingdom. The codes were originally released in 1986 and were intended for use in CPR systems. In 1988 they were adopted as a standard by the National Health Service (NHS) in the U.K. Further development of the codes is being done under the auspices of the NHS Centre for Coding and Classification (NHS CCC). The current version 2 codes are in daily use in general practices throughout the U.K. It is estimated that over 80% of general practices in the U.K. use computerized medical records systems, which is likely the highest rate in the world.[4] The new version 3.1 codes were developed in order to expand the codes to incorporate terms used by hospital-based specialists and other health care personnel and to improve on some of the structural deficiencies present in the earlier versions.

This new version of the codes represents a major restructuring and enhancement of the vocabulary and coding system.[5,6] Some key features of the new structure include: (1) Each code represents a single medical concept. (2) Each concept may be described by multiple terms. (3) The hierarchy is separated from

0195-4210/95/$5.00

the codes themselves. It is now represented by a directed acyclic graph, allowing a term to have multiple parents. More than one hierarchy can be used (e.g. for different specialties). (4) Incorporation of an "information model" for term qualifiers.[5,7] The use of qualifiers allows a richer level of detail in the coding without resulting in an unmanageable expansion of the number of terms. The information model restricts the use of qualifiers so that only the qualifiers that have sensible combinations with a particular term can be used. Another goal of the Read developers is that the thesaurus should contain natural clinical terms, suitable for use by clinicians in describing the process of patient care.[5] A total of 112,835 concept codes and 130,225 terms are included in the July 1994 pre-release version 3.1 distribution, which is roughly comparable to the size of SNOMED International.

It is expected that the significant expansion in the number of terms, and the new hierarchy and qualifier structures should provide a rich terminology for a controlled vocabulary and coding system for use in primary care medicine. Our study is intended to begin to evaluate the applicability of this vocabulary and coding system to the different environment of U.S. primary care medicine.

METHODS

A sample of convenience consisting of all of the dictated progress notes for outpatient clinic visits for one day in March 1995 (a total of 241) was obtained from the University of Missouri (MU) Family Practice clinic. From these notes a random sample of 50 was selected to obtain problem assessment phrases for coding. Three notes were unsuitable due to having no problem assessments recorded, and they were replaced by an additional three randomly selected notes to maintain a total study sample of 50. These notes were recorded by a total of 23 different clinicians, including 7 attending physicians, 10 family practice residents, 2 medical students, 2 nurse practitioners, 1 psychologist, and 1 psychology resident.

From these notes a total of 75 problem assessment phrases were recorded. These phrases were obtained from the "assessment" section of the SOAP note format used at the MU clinic, which often serves as the de facto problem list in addition to recording the problem assessments.[8] Only problem and assessment information was used in this analysis.

A database application was developed using Microsoft Access™ to facilitate recording of the problem assessment phrases, their coded equivalent English terms, and their evaluations. Using the Read code browser application provided by the NHS CCC on the distribution CD-ROM, the principal investigator attempted to find the Read terms which best matched the original phrases. An attempt was made to exhaustively encode the entire phrase, not just the core concept. Multiple codes were in some cases required, and this was noted. Whenever possible, the qualifier terms were used in order to allow for the most detailed coding possible, and the core concept plus all associated qualifiers were considered to be a single code. Conversely, if an appropriate qualifier existed in the vocabulary but was not associated with that particular term, then it was not used, and for the purpose of this analysis was treated as if it did not exist. This is in keeping with the intention of the Read information model, which expects qualifiers to be used only in ways that have been predetermined to be "sensible" and have been incorporated into the vocabulary structure. This is in contrast to some other vocabularies, such as SNOMED, which include qualifier and modifier terms, but have no explicit rules determining how and when they can be used, leaving that to the discretion of the user. A few instances of redundancy in the vocabulary were found, and these were explicitly noted.

The original problem assessment phrases and their encoded representations were then evaluated by the principal investigator. Each problem assessment phrase was evaluated for both complexity and clarity. The complexity of the language of the phrase was evaluated using a 3-point scale, with 1 = "simple," 2 = "intermediate," and 3 = "complex." A simple phrase was defined as consisting of a single concept with one or more associated direct modifiers. An intermediate phrase might include an additional one or more concepts with a specific relation to the main concept, or simply multiple concepts expressed in a single phrase. A complex phrase contains significant additional natural language constructs such as complete or partial sentences and phrases not necessarily modifying a specific concept. The clarity of the clinical meaning of the phrase was evaluated using a 3-point scale, with 1 = "vague," 2 = "intermediate," and 3 = "clear," based on the impression of the evaluator.

The encoded representations of the phrases were evaluated using an acceptability scale and a list of specific types of difficulties encountered during the

coding process. The acceptability scale was used to assess the degree of clinical acceptability of the term as a substitute in the medical record for the original problem assessment phrase, similar to the methods used by Rosenberg and Coultas.[9] A 6-point scale was used, with 0 = "unencodeable," 1 = "extremely dissatisfied," 2 = "dissatisfied," 3 = "neutral," 4 = "satisfied," and 5 = "extremely satisfied."

RESULTS

From the 50 progress notes, a total of 76 assessment phrases were obtained, for an average of 1.5 assessments per note. Only one phrase was duplicated, resulting in 75 unique phrases. Some similar phrases were considered unique if they differed by one or more qualifiers or other words. Spelling variants or word order differences alone would not have been considered unique. However, this did not occur in our sample.

The distributions of the problem assessment phrase characteristic scores are shown in Tables 1 and 2. The mean score for complexity is 1.64, with a standard deviation of 0.777. The mean score for clarity is 2.53, with a standard deviation of 0.573.

Table 1. Problem Assessment Phrase Complexity

Complexity	n (%)
Simple (1)	41 (54.6)
Intermediate (2)	20 (26.7)
Complex (3)	14 (18.7)
Total	75 (100)

Table 2. Problem Assessment Phrase Clarity

Clarity	n (%)
Vague (1)	3 (4.0)
Intermediate (2)	29 (38.7)
Clear (3)	43 (57.3)
Total	75 (100)

A total of 82 Read representations were generated for the 75 problem assessment phrases, with six phrases found to have multiple representations. Three of these redundant representations were inherent in the coding system. These included (1) "ear pain" and "pain in ear," and (2) "individual general health examination," and "general examination of patient," which appeared twice. In both cases, the two phrases were represented by unique concept codes, even though they appear to be indistinguishable clinically.

The other three cases resulted from inability to find a completely appropriate code for a phrase, resulting in the need to select from multiple closely related codes, none of which could be clearly judged to be the best match. One of these resulted from attempting to code the phrase "actinic keratosis left ear." Read codes were found for the main concepts of "actinic keratosis" and "ear." "Ear" includes an allowed qualifier for "laterality," which allows the value "left" to be specified. "Actinic keratosis" includes allowed qualifiers for "due to," "behavior" (the value is specified by the vocabulary as "benign"), "type," and "site." The "site" qualifier has a specified value of "surface anatomy," rather than allowing a choice of specific surface locations on the body, as would be expected. "Surface anatomy," however, allows the additional qualifier of "laterality," which again allows the value "left" to be specified. The resulting term is "actinic keratosis, surface anatomy, left," which unfortunately is far from ideal for clinical terminology. To be complete in this case, the codes for "actinic keratosis" and "ear" must be used together. Which one is given the qualifier "left" is entirely arbitrary, and this results in redundancy. The redundancy in this case would be eliminated if the information model for "actinic keratosis" was complete.

The distribution of the acceptability scores for the coded Read terms is shown in Table 3. The mean score is 3.17, with a standard deviation of 1.18.

Table 3. Read Term Acceptability

Acceptability	n (%)
Extremely Satisfied (5)	14 (17.1)
Satisfied (4)	15 (18.3)
Neutral (3)	31 (37.8)
Dissatisfied (2)	16 (19.5)
Extremely Dissatisfied (1)	5 (6.1)
Unencodeable (0)	1 (1.2)
Total	82 (100)

Table 4 shows the distribution of the number of Read codes required to represent a single problem assessment phrase. The mean number of codes required is 1.57, with a standard deviation of 0.88.

Table 4. Number of Read Codes Required for Problem Assessment Phrase Representation

Number of Codes Required	n (%)
0	1 (1.2)
1	49 (59.8)
2	21 (25.6)
3	6 (7.3)
4	5 (6.1)
Total	82 (100)

From the total of 82 problem assessment phrase representations, 64 (78%) were found to have one or more coding difficulties, with a total of 122 instances noted. After eliminating multiple entries of the same difficulty type for a specific phrase, due to redundant representations, a total of 111 coding difficulties remained, resulting in an average of 1.48 per phrase (n = 75). A total of 13 different types of coding difficulties were noted. Multiple difficulty types could be present in a problem assessment phrase representation. The "% of phrases" column in Table 5 records the percentage of the total number of problem assessment phrases (n = 75) in which the particular type of coding difficulty was noted.

Table 5. Coding Difficulties

Coding Difficulty	n (%)	% of phrases
1. code too specific	8 (7.4)	10.7
2. code too general	1 (0.8)	1.3
3. qualifier term not present	15 (12.3)	20
4. qualifier term present but not allowable for that Read concept	18 (18.0)	24
5. British spelling variant	6 (5.7)	8
6. language not clinically appropriate	8 (8.2)	10.7
7. lack of appropriate conjunction, linking term or relation	10 (8.2)	13.3
8. British abbreviation variant	2 (1.6)	2.7
9. phrase too complex	9 (7.4)	12
10. coded term has significantly different wording	15 (12.3)	20
11. element of phrase not codeable	11 (10.7)	14.7
12. multiple codes required for single concept	7 (6.6)	9.3
13. term in wrong level of hierarchy	1 (0.8)	1.3
Total	111 (100)	

DISCUSSION

The majority of the problem assessment phrases were rated as "simple" (54.6%) and "clear" (57.3%). However, a substantial minority of the phrases remained which were judged as more complex (45.4%) and less clear in meaning (42.7%). It is important to consider the characteristics of the natural language phrases normally used by clinicians when attempting to develop or implement a standardized vocabulary to be used as a substitute. Further work is needed to better characterize the phrases used by the clinicians and to investigate in greater detail how these characteristics impact the process of coding.

Redundancy should continue to be addressed in further coding system development, in an effort to identify and eliminate or explicitly deal with truly redundant concepts.[10] If redundancy is recognized and dealt with in a manner transparent to the user, then it ceases to be a problem. Apparent redundancy that occurs due to lack of appropriately specific terms for coding can probably only be addressed by continuing vocabulary expansion and maintenance.

The acceptability scores for the Read representations show that 37.8% were rated as 3 (neutral). Only 35.4% were rated as 4 or 5, while 26.8% were rated as 0 to 2. These ratings are significantly lower than those reported by Rosenberg and Coultas,[9] who found that 58.5% of the UMLS terms were rated as 5 (extremely satisfied). However, they did note significant intrarater variability, with a mean correlation of .75, compared to a mean intrarater correlation of .94. Further assessment is needed to determine the relative contributions of the evaluator and terminology differences to these results.

A total of 59.8% of the problem assessment phrases were represented with a single code, with 25.6% requiring 2, and 13.4% requiring 3 or 4. In addition, the coding difficulties list shows a lack of appropriate conjunctions, linking terms, or relations in 13.3% of cases. No explicit mechanisms are available to link multiple codes in a standardized way in Read or any of the other controlled clinical vocabularies, which suggests the need to investigate more robust models of medical concept representation. One possibility would be the conceptual graphs model, as has been proposed with SNOMED codes.[11]

Two of the three most common coding difficulties were (1) qualifier terms that were not present in the vocabulary, or (2) were present but could not be used for a particular concept, even when they were entirely appropriate. The information model for qualifiers needs to be expanded to allow clinicians to be able to use all sensible combinations of qualifiers with a given core term. This would significantly reduce the need to code concepts using multiple codes, further reducing the possibility of redundancy. Much additional development will be required, but this should result in significant benefits in both coding

and retrieval once adequate completeness has been achieved. It is expected that the next release of the codes will have significantly expanded qualifier content and structure.[5] An alternative approach could be to remove the restrictions on qualifier usage, leaving it entirely to the user's discretion, as in SNOMED. This would significantly reduce the development effort required, and would eliminate the possibility of existing and appropriate qualifiers being unavailable. However, it would also increase the burden on the clinician in searching for and selecting appropriate terms and qualifiers from a complex terminology at the time of clinical documentation, which may be unacceptable. It would likely also result in less standardization and increased redundancy in the resulting representations.

A concern with using the Read vocabulary in U.S. practice is the possibility of significant language variation. This was not noted to be a significant problem in this study, with both spelling and abbreviation variants affecting the coding only in a total of 10.7% of cases. Most of these variations were minor, as in "dysmenorrhoea" and "haemorrhoids," or "URTI" instead of "URI" for acute upper respiratory infection. This could easily be dealt with by adding synonymous terms for these well-recognized variations.

There are several limitations which must be considered in interpreting the results of this study. One is the relatively small sample size. We plan to expand the sample for future work, and this will allow for more detailed statistical analysis. Another limitation is the potential bias introduced by using a sample of convenience. This did not appear to be a significant problem, at least partially due to the fairly large number of clinicians represented, and the variety of diagnoses that were present in the sample. A major limitation in the study is the use of a single coder and evaluator. This could result in significant bias in the evaluation, and may account at least partially for the relatively low term acceptability scores. Future plans include recruiting a group of clinicians to independently evaluate both the problem assessment phrases and their coded representations.

CONCLUSIONS

The Read Codes appear to be one of the best potential candidates for adoption as a standard vocabulary for use in CPR systems. This appears to be true in the U.S., as well as in the U.K. The modifications necessary for the terminology to be clinically useful in the U.S. appear to be fairly minor. Further development work is needed, however, in order to expand the content and structure of the information model for qualifiers.

Acknowledgment

This work was supported in part by training grant LM-07089 from the National Library of Medicine.

References

1. Dick RS, Steen E. The computer-based patient record: an essential technology for health care. Washington (DC): National Academy Press, 1991.
2. Bernstein RM, Hollingworth GR, Viner G, Lemelin J. Family practice informatics: research issues in computerized medical records. Proceedings of the Seventeenth Annual Symposium on Computer Applications in Medical Care; 1993:93-7.
3. Vocabularies for computer-based patient records: identifying candidates for large scale testing (minutes of meeting 5-6 Dec 1994). Washington, (DC): National Library of Medicine and AHCPR, 1994.
4. Read Codes and the terms projects: a brief guide. Woodgate, Leicestershire, UK: NHS Centre for Coding and Classification, 1994.
5. O'Neil M, Payne C, Read J. Read Codes Version 3: a user led terminology. Woodgate, Leicestershire, UK: NHS Centre for Coding and Classification, 1994.
6. Read Codes file structure version 3: overview and technical description. Woodgate, Leicestershire, UK: NHS Centre for Coding and Classification, 1993.
7. A proposed new file structure for qualifiers. Woodgate, Leicestershire, UK: NHS Centre for Coding and Classification, 1993.
8. Weed LL. Medical records that guide and teach. N Engl J Med 1968; 278:593-600.
9. Rosenburg KM, Coultas DB. Acceptability of Unified Medical Language System terms as substitute for natural language general medicine clinic diagnoses. JAMIA Proceedings of the Eighteenth Annual Symposium on Computer Applications in Medical Care; 1994:193-7.
10. Cimino JJ, Clayton PD, Hripcsak G, Johnson SB. Knowledge-based approaches to the maintenance of a large controlled medical terminology. JAMIA 1994;1:35-50.
11. Campbell KE, Das AK, Musen MA. A logical foundation for the representation of clinical data. JAMIA 1994;1:218-32.